CLAIMS

What is claimed is:

1. An isolated oligonucleotide comprising at least one ditag, wherein the ditag includes two joined first and second sequence tags, and wherein the first tag includes a 5'-terminus sequence and a second tag comprises the 3'-terminus sequence of a nucleic acid molecule or fragment thereof.

2. The oligonucleotide of claim 1, further comprising two adapters flanking the ditag, wherein each adapter includes at least one restriction site.

3. The oligonucleotide of claim 2, wherein each adapter comprises at least a first restriction site which is an asymmetric restriction site and at least a second restriction site.

4. The oligonucleotide of claim 3, wherein the asymmetric restriction site comprises a type II restriction site.

5. The oligonucleotide of claim 1, wherein the nucleic acid molecule comprises the full-length sequence of a gene or a fragment thereof.

6. The oligonucleotide of claim 1, wherein the nucleic acid comprises RNA, mRNA, genomic DNA, full-length cDNA, or cDNA.

7. The oligonucleotide of claim 1, wherein the ditag is obtained by splicing the 5' terminus and the 3' terminus of the nucleic acid molecule or fragment thereof in the presence of at least one restriction enzyme and the size of the sequence tags is determined by the restriction enzyme used.

8. The oligonucleotide of claim 7, wherein the restriction enzyme is a type II restriction enzyme.

9. The oligonucleotide of claim 7, wherein the restriction enzyme is selected from the group consisting of AarI, AceIII, AloI, BaeI, Bbr7I, BbvI, BbvII, BccI, Bce83I, BceAI, BcefI, BcgI, BciVI, BfiI, BinI, BplI, BsaXI, BscAI, BseMII, BseRI, BsgI, BsmI, BsmAI, BsmFI, Bsp24I, BspCNI, BspMI, BsrI, BsrDI, BstF5I, BtgZI, BtsI, CjeI, CjePI, EciI, Eco31I, Eco57I, Eco57MI, Esp3I, FalI, FauI, FokI, GsuI, HaeIV, HgaI, Hin4I,

HphI, HpyAV, Ksp632I, MboII, MlyI, MmeI, MnlI, PleI, PpiI, PsrI, RleAI, SapI, SfaNI, SspD5I, Sth132I, StsI, TaqII, TspDTI, TspGWI, TspRI, Tth111II, I-CeuI, PI-SceI, PI-PspI and I-SceI.

10. The oligonucleotide of claim 1, wherein the ditag size is 12-60 bp.

11. The oligonucleotide of claim 1, wherein the ditag comprises 34-38 nucleotides and the size of each tag is determined by the use of restriction enzyme MmeI.

12. The oligonucleotide of claim 1, wherein the first and second tag have the same or a different number of nucleotides.

13. The oligonucleotide of claim 1, wherein the oligonucleotide consists of 1 to 1000 ditags.

14. A vector comprising an isolated oligonucleotide including at least one ditag, wherein the ditag includes two joined first and second sequence tags, and wherein the first tag includes a 5'-terminus sequence and a second tag comprises the 3'-terminus sequence of a nucleic acid molecule or fragment thereof.

15. The vector of Claim 14 further comprising two adapters flanking the ditag, wherein each adapter includes at least one restriction site and wherein each adapter includes at least a first restriction site which is an asymmetric restriction site and at least a second restriction site.

16. The vector of claim 15, wherein the backbone of the vector does not comprise the asymmetric restriction site or the second restriction site.

17. The vector of claim 16, wherein the asymmetric restriction site is a type II restriction site.

18. A vector comprising at least a nucleic acid molecule and two adapters flanking the nucleic acid molecule, wherein each adapter comprises at least: a first restriction site which is a type II restriction site and at least a second restriction site, and wherein the backbone of the vector does not comprise the type II restriction site, or the second restriction site.

19. The vector of claim 18, wherein the type II restriction site is MmeI.

20. A vector comprising SEQ ID NO:18.

21. The vector of claim 20, comprising at least one ditag, wherein the ditag includes two joined first and second sequence tags, and wherein the first tag includes a 5'-terminus sequence and a second tag comprises the 3'-terminus sequence of a nucleic acid molecule or fragment thereof.

22. The vector of claim 21, comprising comprising two adapters flanking the ditag, wherein each adapter includes at least one restriction site and wherein each adapter includes at least a first restriction site which is an asymmetric restriction site and at least a second restriction site.

23. A cDNA library, wherein every cDNA clone of the library comprises at least one oligonucleotide including at least one ditag, wherein the ditag includes two joined first and second sequence tags, and wherein the first tag includes a 5'-terminus sequence and a second tag comprises the 3'-terminus sequence of a nucleic acid molecule or fragment thereof.

24. The cDNA library of claim 23, wherein the at least one oligonucleotide comprises 1-1000 ditags.

25. A method for preparing at least one oligonucleotide including at least one ditag comprising:

producing at least one nucleic acid molecule;

isolating the 5' terminus and the 3' terminus of the nucleic acid molecule or fragment thereof; and

linking the 5' terminus and 3' terminus to create the at least one ditag.

26. A method for preparing at least one oligonucleotide comprising at least one ditag comprising:

providing at least one nucleic acid molecule flanked by two adapters;

isolating the 5' terminus and the 3' terminus of the nucleic acid molecule; and

linking the 5' terminus and 3' terminus to create at least one oligonucleotide including at least one ditag flanked by the two adapters.

27. The method of claim 26, further comprising the step of creating concatemer of ditags.

28. The method of claim 27, wherein the ditags are 1-1000 ditags.

29. The method of claim 26, further comprising including the oligonucleotide comprising the at least one ditag flanked by the adapters in a vector.

30. The method of claim 26, wherein the nucleic acid molecule is RNA, mRNA, genomic DNA, full-length cDNA, or cDNA.

31. The method of claim 26, further comprising the step of determining the nucleotide sequence of the at least one ditag to detect gene expression.

32. The method of claim 26, further comprising:

determining the sequence of the at least one ditag; and

comparing the ditag nucleotide sequence to a database comprising genomic sequences whereby matching 5' and 3' termini sequences are identified.

33. The method of claim 26, comprising:

producing at least one nucleic acid molecule flanked by two adapters, wherein each adapter comprises at least one restriction site;

splicing the 5' terminus and the 3' terminus of the molecule to produce at least one ditag, wherein splicing includes adding at least one restriction enzyme capable of recognizing the recognition sites.

34. The method of claim 33, wherein the two recognition sites are asymmetric recognition sites.

35. The method of claim 34, wherein the asymmetric recognition sites are restriction endonuclease asymmetric cleavage site sequences recognizable by type II restriction enzymes, and wherein the type II restriction enzyme is selected from the group consisting of AarI, AceIII, AloI, BaeI, Bbr7I, BbvI, BbvII, BccI, Bce83I, BceAI, BcefI, BcgI, BciVI, BfiI, BinI, BplI, BsaXI, BscAI, BseMII, BseRI, BsgI, BsmI, BsmAI, BsmFI, Bsp24I, BspCNI, BspMI, BsrI, BsrDI, BstF5I, BtgZI, BtsI, CjeI, CjePI, EciI, Eco31I, Eco57I, Eco57MI, Esp3I, FalI, FauI, FokI, GsuI, HaeIV, HgaI, Hin4I, HphI, HpyAV, Ksp632I, MboII, MlyI, MmeI, MnlI, PleI, PpiI, PsrI, RleAI, SapI, SfaNI, SspD5I, Sth132I, StsI, TaqII, TspDTI, TspGWI, TspRI and Tth111II

36. The method of claim 34, wherein the asymmetric recognition sites are homing endonuclease asymmetric recognition site sequences, and wherein the enzyme recognizing the homing endonuclease asymmetric restriction site is selected from the group consisting of : I-CeuI, PI-SceI, PI-PspI and I-SceI.

37. The method of claim 26, comprising:

producing at least one full-length cDNA molecule comprising two adapters flanking the 5' terminus and 3' terminus of the full-length cDNA, wherein each adapter comprises at least one restriction site which is a MmeI recognition site ;

splicing the 5' terminus and the 3' terminus of the full-length cDNA to produce at least one oligonucleotide comprising at least one ditag, wherein splicing includes cleaving the full-length cDNA with MmeI which forms 3' overhanging tag ends, and ligating the two 5' and 3' termini tags to produce the ditag.

38. The method of claim 26, wherein the wherein the at least one ditag comprises 34-38 nucleotides.

39. A method for genome mapping, comprising:

preparing at least one oligonucleotide including at least one ditag, the ditag including two joined first and second sequence tags, wherein the first tag includes the 5'-terminus sequence and the second tag includes the 3'-terminus sequence of a nucleic acid molecule, the nucleic acid molecule corresponding to the full-length of a gene or fragment thereof;

mapping each of the two tags of the at least one ditag on the genome; and

defining the structural region of the corresponding gene on the genome map.

40. A method of gene discovery comprising:

preparing at least one oligonucleotide comprising at least one ditag, the ditag including two joined first and second sequence tags, wherein the first tag includes the 5'-terminus sequence and the second tag includes the 3'-terminus sequence of a nucleic acid molecule, the nucleic acid molecule corresponding to the full-length of a gene or fragment thereof;

comparing the obtained at least one ditag with a genome map and/or a gene database;

detecting matching of the 5' and 3' termini tags on the genome map but detecting no match on one ore morer gene database.

41. The method of claim 40, further comprising recovering the full-length nucleic acid molecule corresponding to the newly discovered gene.

42. A method for recovering full-length cDNA comprising:

preparing, from a full-length cDNA library, at least one oligonucleotide including at least one ditag, the ditag including two joined first and second sequence tags, wherein the first tag includes the 5'-terminus sequence and the second tag includes the 3'-terminus sequence of a full-length cDNA;

sequencing the obtained oligonucleotide ditag;

determining the ditag of interest; and

recovering the full-length cDNA corresponding to the ditag of interest from the full-length cDNA library.

43. A method for quantifying the transcriptional activity of a gene comprising:

preparing, from a full-length cDNA library, at least one oligonucleotide comprising at least one ditag, the ditag including two joined first and second sequence tags, wherein the first tag includes the 5'-terminus sequence and the second tag includes the 3'-terminus sequence of a full-length cDNA;

sequencing the obtained oligonucleotide ditag;

determining the frequency of the sequenced ditag which corresponds to the transcriptional activity of the gene.